Legal Issues in AI Forensics: Understanding the Importance of Humanware

# Legal Issues in AI forensics: understanding the importance of humanware

**Raffaella Brighi**[1] , **Michele Ferrazzano**[2] , **Leonardo Summa**[3]
[1] Associate Professor of Digital Forensics, Alma Human AI - Department of Legal Studies, University of Bologna
[2] Adjunct professor of Information technology and Research fellow - Department of Legal Studies, University of Modena and Reggio Emilia
[3] Alma Human AI - Department of Legal Studies, University of Bologna
raffaella.brighi@unibo.it, michele.ferrazzano@unimore.it, leonardo.summa@studio.unibo.it

**Abstract**

Nowadays, our most cogent need is to embrace a new vision of the digital forensics field, which requires to be focused on: (a) the harmonization of the legal framework and technical standards; (b) the pursuit of common paths when conducting forensic investigations; and (c) the definition of an epistemological frame of reference. These three elements should be intended as the cornerstone of this change. The growing influence that ICT technology is having on the work of judges and legal professionals now requires a stronger holistic basis—concerning principles, practices, and procedures—of what is available, namely, *humanware*, and what is useful, namely, AI, to achieve and disseminate best practices. Firstly, the full potential of AI calls for a deep insight into its technical implications and into the requirements needed to keep operating in a forensic-based environment, but it also calls for deep understanding by policymakers, who may lack a sense for the ethical and legal implications of AI, while pushing for its deregulation. Therefore, understanding the urgency to act for the development of a strong and well-trained *humanware* is just the baseline in tackling well-known problems in the application of AI technologies (e.g. the reliability and explainability of machine learning methods) in the digital forensics field, as well as in the whole of society.

## 1 Introduction

In recent times, a debate has been ignited in the juridical world endeavouring to regulate the deployment and the possible applications of artificial intelligence (AI). Having legal decisions supported by AI is an appealing idea that dates back several years (Sartor 1992; Sartor 1998).

Numerous expert systems have been developed in the past, with the aim of autonomously reaching decisions by exploiting the representation of specialized legal knowledge in symbolic form, with logical rules and predefined inferences: the outcomes, however, were less promising than expected. Nevertheless, AI has evolved, owing to highly effective machine learning methods that deploy the knowledge deriving from big data analysis (Russell and Norvig 2009). Consequently, these recent developments have questioned both the introduction of AI technologies in different legal systems and its ethical-legal sustainability be questioned (Floridi et al. 2018).

An evidence of the great potential of these tools can be found in the widespread application of intelligent agents in support of daily and repetitive actions. At the same time, it indicates that the legal consequences of an unregulated use should be taken into account and prevented (Lasagni and Contissa 2020).

Examples include the potential probative interest profiles guarded by intelligent devices, a subject studied by IoT forensics, predictive capabilities and the fallacious discriminatory bias, the effectiveness and usefulness of the results obtained in terms of reliability and, finally, the remedies we should choose in overcoming the limits that have become apparent (Sommaggio and Marchiori 2020).

There is now widespread news, as well as numerous studies, concerning robot-judges (Millar and Kerr 2013), AI systems in a position to predict the potential criminal activities (i.e. so-called predictive policing), or even algorithms assessing an individual's social dangerousness, such as the COMPAS system implemented in U.S. courts to quantify the risk of recidivism, within the frame of predictive justice (Degeling and Berendt 2018). However, although these applications are already in the testing phase, the full potential of AI might be underestimated.

The use of AI in the collection and forensic analysis of digital evidence could be the real breakthrough that can help the justice system to streamline procedures, primarily by shortening the timeframe of investigations. It is evident that digital forensics (DF) faces mounting challenges in terms of accuracy and timeliness in the analysis of a growing amount of data from increasingly diverse sources (Council and others 2009).

Thus, a question arises as to what application of AI may effectively optimize investigation time and ensure the reliability of the results of digital evidence analysis. The aim of the present paper is to answer this question by investigating the sustainable and desirable points of contact between AI applications and the substantive and procedural rules to be observed during investigation activities, though they might differ from the traditional forms.

The keys to a productive dialogue lie in the human factor, in forensic IT experts acquiring sufficient knowledge of these tools, and in legal practitioners becoming sensitized to forensic IT issues (Brighi and Maioli 2016). If AI applications in digital forensics are to be properly regulated, their operating mechanisms need to be fully comprehended, and the boundaries between legally acceptable and unacceptable

consequences must be set, rather than enthusiastically embracing uptake at all costs and shifting the burden of damages to end users, both in the legal area and in our daily interaction with these technologies (Abdelnasser Gamal 2020).
The future of AI is clear now. The challenge is to have these instruments formally accepted in court proceedings by grounding their use in fundamental rights and fair trial principles. This work aims to endorse the role of the human factor in the sedimentation of today's digital transformation by highlighting the friction generated with the legal categories of reference and fostering the development of skills and tools by which to manage such promising technologies. The *raison d'être* of this work is indeed the human-based vision of the coexistence of our modern society with new technologies, rooted in the neutrality of the latter and the fertile *Weltanschauung* that has allowed the development of such revolutionary tools.
We aim to identify the legal issues arising in connection with the adoption of AI by DF and to suggest possible solutions. Section 2 provides an overview of state-of-the-art AI applications in DF investigations and highlights the constraints and benefits that can be derived from their implementation.
Section 3 analyses the legal consequences, in terms of compression of the right of defence, violation of the legal principles protecting the fundamental rights of individuals, quantification of the acceptable margin of error regarding findings of guilt, solidity in terms of the logicality and coherence, and verifiability of results capable of satisfying the obligation to justify judicial measures adopted in cooperation, in whole or in part, with AI-based instruments.
Finally, Section 4 illustrates some future guidelines to be followed for the construction of a desirable synergy between techniques and law, between humanware and progress in the field of ICT.

## 2 The importance of AI into the Digital Forensics field

The last decade has witnessed the conversion of most data, such as books, videos, pictures, and medical information, into digital formats. Laptops, tablets, smartphones, and wearable devices are the major enablers of this digital data transformation and have become a substantial part of our daily lives.
As a result, we are becoming a soft target for many forms of cybercrimes. Digital forensic investigation seeks to recover lost or deliberately deleted or hidden files from a suspect's device. However, due to underdeveloped skills and lack of time, current human capabilities and government resources are insufficient for cybercrime investigations.
Existing digital investigation procedures and practices require time-consuming human interactions, thus slowing down the entire process. Many research projects, studies, and even some professional products have begun to offer solutions based on artificial intelligence to overcome known obstacles.
However, a focus on what AI is would take us away from the purview of this work. Different approaches have been tried in the history of AI which have variously paid attention to the mental models and human reasoning or to human behaviour, in attempt to develop systems that simulate human tasks execution and to build either ideally intelligent systems or systems that employ rational behaviours in order to act properly. For the purposes of our paper, AI can be considered as an instrument capable of conducting and facilitating human tasks.
AI technology is growing day by day, and its widespread use increases the number of malicious activities, with some relevant issues arising about their legal attribution (King et al. 2020). Artificial intelligence programs are called intelligent agents, and they are used to interact with the environment. The agent uses different techniques to identify the environments through its sensors, and then it can take the action needed to achieve the desired state through its sensors. The important aspects in AI technologies are how the sensors are used to collect data and how they map them onto the actuators; this is how the functions within agents can achieve these results.
A rational agent does not limit itself to gathering information but must be able to learn as much as possible by accumulating experience. Machine learning (ML) is a specific part of artificial intelligence that enables computers to learn without being explicitly programmed. For example, a machine learning system is able to find patterns in data and use them to predict the outcome of something it has never seen before.
AI technologies afford significant advantages and have a bright future ahead. However, these technologies are also unavoidably used to carry out some serious crimes that can be dangerous for people (King et al. 2020; Ferrazzano 2019).
Below is an overview of AI applications in DF investigations, highlighting constraints and benefits.

### 2.1 ML/AI & Incident Response

Until recently, cyberattacks were dealt with by relying on basic antivirus software or firewall with a list of rules. However, current cyberattacks are sophisticated enough to bypass traditional security measures. This is owed to limited human expertise and efficiency, which in turn can be attributed to several causes: the time required to detect and investigate daily threats, lack of skills, lack of accuracy, failure to detect advanced threats such as advanced persistent threats (APTs), ransomware, or fileless attacks (Ghafir et al. 2018).
AI can efficiently handle cybersecurity threats by rapidly detecting and analysing millions of logs and anomalous events, identifying a malicious file, or recognizing an atypical behaviour from a seemingly harmless data cluster or file. Security strategists can provide current advanced machine learning models with a massive quantity of historical training data, achieving increasingly better security responses when more valuable data are provided.
A practical example that displays who and what could benefit from machine learning is represented by the Security Operations Centers (SOCs). A SOC is a facility that hosts an information security team responsible for continuously monitoring and analysing an organization's security posture: the goal is to detect, analyse, and respond to cybersecurity incidents by using a combination of technology solutions and a strong set of processes. Given the number of sources of rel-

evant data alone, the impracticality of manually reviewing log files is apparent.
This challenging obstacle is traditionally overcome by relying on a system that correlates inputs by dozens of different security products, each monitoring a specific attack vector, so as to notify the SOC about the occurrence of an unusual event.
Since the SOC writes these correlation rules after the occurrence of an incident – in order to be notified of its reoccurrence – there are two main downsides. Firstly, several important events are missed because correlation rules rely on a specific set of inputs. If excessively narrow rules are defined by the SOC, the system will not be triggered by minimally different events. Considering the intra-organization variability in applications, systems, and environments, it is unlikely that two attacks will be identical. Secondly, false positive results can be generated if the rules are not narrow enough: this poses the risk of masking real attacks by generating countless alerts that cannot be readily filtered by the SOC to identify real threats.
Either way, analysts miss attacks in the deluge of data, or they identify them too late. In order to find important security events without generating low value alerts that demand time, attention, and manual remedy, the SOC may leverage AI and ML.
Let us recall that AI is a broad term that refers to algorithms, models, and a field of scientific study. ML is the concept of training a system to perform narrowly focused tasks without using explicit instructions, relying on pattern detection and conclusion inference. It focuses on a specific need.
AI and ML can identify important security events in an organization, with high accuracy, by gathering together data from multiple sources while optimizing the time and experience required in the SOC. To date, many security companies have developed products that work with ML algorithms to try to help companies fight cybercrime [1] [2] (Trifonov et al. 2019; Hasan et al. 2011).

## 2.2 ML/AI & Forensics Analysis and Evaluation

An increasingly important area in computing, digital forensics frequently requires the intelligent analysis of large amounts of complex data: most challenges currently posed by these needs may be ideally approached through AI. An important issue for AI in the forensic arena is the ability to explain the reasoning process (Krivchenkov, Misnevs, and Pavlyuk 2019).
Two subtypes of AI techniques are recognized: symbolic (techniques reasoning with discrete entities in a knowledge base) and sub-symbolic (techniques in which the knowledge is spread across the representation structure). Expert systems are a common example of symbolic AI techniques: they follow a predefined rule base, and normally rely on a regulated strategy to select which rule to use at any particular moment in time.
Therefore, expert systems can, at any point, provide an explanation of the reasoning for the conclusions obtained, thus permitting an outside entity to review the reasoning process and to recognise any flaws in the reasoning itself (Mitchell 2014).
However, two major drawbacks of symbolic systems can be identified. The first of these drawbacks is that they operate in a closed world: any item that is not part of the rule base cannot be Justified in the reasoning process.
This is a serious issue in a rapidly evolving area such as computing, as rebuilding a rule base *de novo* is a time-consuming task and adding additional rules (a process known as "rule base repair") can damage the original performance.
The second drawback is that expert systems perform poorly with large quantities of data. This is a major disadvantage in digital forensic investigations, where exponentially larger amounts of data need to be investigated. However, techniques such as expert systems might prove to be useful in higher-order situations, such as suggesting the following steps to an investigator, or advising on what an organisation's policy should prefer in a given situation (Costantini, De Gasperis, and Olivieri 2019).
A form of typically symbolic AI that may bypass the disadvantages of expert systems (and other symbolic rule-based systems) is that of case-based reasoners (CBRs). CBRs are built on psychological notions concerning information representation by domain experts themselves.
Most domain experts heavily rely on their past experiences: when faced with an issue, they will draw parallels between current and past situations, thus using first principles to find a solution only when all possible similar cases in their experience have been exhausted. Similarly, a CBR system first collects a large number of cases (and, in digital forensics, the resulting actions), and then resorts to a metric to relate the current situation to one already included in the case base. If a perfect match is found, then the current situation will be managed through the same solution applied in the initial case.
Likewise, if a partially similar match is found, the system may attempt to adapt the action of the matched case to the current situation employing the so called "repair" rules. CBR systems have the advantage of approaching a problem in a way that is familiar to the expert, while coping with large amounts of data, and dealing with entirely unknown situation.
Since the reasoning can be inspected (this case was closest to X, and in X you did Y), CBR system also expose their reasoning process. Consequently, the quality of the cases and the number of different scenarios included in the case base are crucial to determine the performance of CRBs. A further limit of CRBs is that, while they can support the investigation, they might be ill-suited to lower-level activities (i.e. "find all pictures with naked people in them") (Sanchez et al. 2019).

[1] Microsoft uses its own cybersecurity platform, Windows Defender Advanced Threat Protection (ATP), for preventative protection, breach detection, automated investigation and response.

[2] Splunk software has a variety of applications, including IT operations, analytics and cybersecurity. It's designed to identify a client's current digital weak points, automate breach investigations and respond to malware attacks. Products like Splunk Enterprise Security and Splunk User Behavior Analytics use machine learning to detect threats so they can be quickly eliminated

Identifying specific types or clusters of data in an investigation is best handled by a type of AI known as "pattern recognition". The type of pattern recognition that people are most familiar with is perhaps image recognition, where software attempts to identify parts of a picture.
Furthermore, there are many other examples of pattern and image recognition, such as detecting a pattern in a SPAM e-mail, or a pattern in a disk image that might indicate it is part of a sound file. Many of the techniques used rely very heavily on statistics or probabilistic reasoning, or both.
The most complex and accurate forms of image recognition that can be used to locate certain types of picture, rely on the awareness of how human perceptual system works. However, at these tools currently have a high rate of false positives and false negatives (depending on where the thresholds are set), besides being very computationally intensive.

## 3 Legal and Ethical issues

The relationship between technology and the law recalls the second of Zeno's four paradoxes of movement, that of Achilles and the tortoise. According to this paradox Achilles, representing the law, races against but will never be able to overtake the tortoise, representing technology.
In this endless chase, the law has often tried to model the existing concepts whenever the relevant transformations produced by computer osmosis in legal relations have generated distortions that are no longer tolerable for the legal system itself. Consequently, reinforced protection at European level has become necessary to regulate the processing of personal data. Similarly, we argue that it is necessary to develop a regulatory framework for the investigative uses of technology that guarantees respect for procedural principles and the fundamental rights of individuals.
For this to materialise, it is necessary to become involved in the constant development and updating of computer skills useful for the construction of investigative models that comply with fundamental rights. This is what we call *humanware*, referring to the human factor that intervenes in digital investigations as well as in the relationship with technology.
Focusing on the growth of a more conscious *humanware* by encouraging certified training course for DF examiners, lawyers and judges will limit the potential pathogenic causes – such as discrimination and bias, margins of error, false positives, false negatives – of unlawful decisions based on AI system. Thus, it will be possible to achieve greater respect for fundamental rights, regarding the application of AI-based systems.
In this section, we will examine the repercussions in terms of the substantive and procedural rights generated by the application of AI tools in the formation of digital evidence, with particular attention to the principles that distinguish civil law with an adversarial legal system.

### 3.1 *Male captum bene retentum*

The legal issue around the usability of illegally acquired evidence is of extreme relevance and known in every legal system. The legal dispute involves a very important question: can testimony constitute fully usable evidence when obtained by illegal means, such as torture?
In this extreme context, two opposing factions can be distinguished: those who claim that such results are also illegitimate—*the fruit of the poisonous tree doctrine*—and those who, on the contrary, save the evidentiary results in the light of the Latin principle of *Male captum bene retentum*.
The rationale behind this latter principle is to safeguard the results of investigations, even if they are achieved through the violation of those procedural rules that protect the fundamental rights of persons subject to judicial ruling.
This theory expresses the problematic synthesis of two opposing requirements that are difficult to reconcile: on the one hand, the need to ensure sources of evidence even by using instruments not typified by procedural rules and, on the other hand, the need to safeguard the guarantees put in place to protect against abuses and violations of internationally recognized fundamental rights. The legal ethical sustainability of AI applications in the DF field cannot prescind from the analysis of this contradiction (Losavio et al. 2019; Abdelnasser Gamal 2020).
Accordingly, it is essential to be aware of the legal effects of the use of such technologies, which cannot accept silent adaptations and advocate the greatest possible sharing in the definition of the criteria, limits, and benefits deriving from the introduction of such technologies into the legal arena. Such a phase transition, with the legal implications of these instruments being carefully assessed, is paramount, lest the function of social defence of the law be transmuted into a contractual relationship supported by the mere criteria of efficiency and usefulness unrelated to its social function (Sanger 2018).
In other words, without such a phase transition, the procedural position of each of us would become as a stock exchange listing, fuelled by the logic of reducing the workload of the courts and ensuring greater efficiency, in comparison with human judgment. And it is precisely in contrast to such a logic that we will have to construct proceedings-sustainable variations of the different AI applications available in the field of digital evidence.
Technological transformation must be reconciled with respect for the fundamental rights of the individual, around which the boundaries of law are drawn: the right to a fair trial, which incorporates the right to an impartial judge; the presumption of innocence until otherwise proven, and the duty of judicial authorities to give reasons for their ruling (Vuille, Lupària, and Taroni 2017).
The question appears Hamletic: how can the need to make judicial processes efficient coexist with thee respect for procedural rules and individual fundamental rights?
The answer is to be found in a more mature symbiosis than the one we are currently experiencing, guided by people's awareness of the instruments, both in sustaining their usefulness and in paying attention to its pathological evolutions.
Public debate should be encouraged to become aware of the legal conscience, which is now weak, in order to raise and stimulate active participation in the formation of judicial practices, while respecting the fundamental rights rec-

ognized at the international level (Quattrocolo et al. 2020).
The first step is to realize the biunivocal character that marks the relationship between *technè* and law, by arising a section in the criminal and civil procedure code dedicated to computer investigations and digital evidence acquisition processes. Specific guidelines and procedures must be provided to ensure compliance with the technical principles of digital forensics and fundamental human rights.

### 3.2 Beyond a reasonable doubt

When assessing the sustainability of the use of AI-based systems in the DF field, another consideration might arise: the introduction of AI-based technologies into evidence generation is strongly conditioned by the degree of reliability achievable in the design of such systems.
The provocative tone of the question offers an opportunity to reflect on the function of these technologies in legal systems. When using AI-based techniques (ANNs, K-means, NLP, etc.), the result that is obtained is reliable by the measure of the margin of error known for that particular system. The acceptable range of error for a given legal system is to be defined in the same way as the degree of tolerance within which human error is justified (Kotsoglou 2019).
The matter of transparency and justifiability of the choices and results produced is a well-known technical problem and cannot be underestimated when applying AI to legal reasoning. Eliminating the risk attendant on the factors of human error (i.e. prejudices, likes/dislikes, personal beliefs, emotional distress) and their consequent influence on the decision-making process is an appealing concept. However, we eventually accept decisions that are unquestionable because the original mechanism producing the result is unexplored (Grace 2019).
For instance, a crucial aspect of paedopornographic crimes is age determination of the victims. The automation of the processes of identification and attribution of the underage factor would be of extraordinary value (Anda, Le-Khac, and Scanlon 2020).
Nevertheless, attention should be paid to some basic considerations:

- Dataset training: checking the input that data used to train neural networks is fundamental. The chosen model is initially built around a training dataset which is a set of examples used to set parameters for the model (e.g., skin tone, height, etc.). To evaluate whether a model is being trained correctly, it is necessary to take note of the loss: the smaller the loss, the better a model. The loss is calculated on the basis of training and validation and can be interpreted by how well the model is doing for these two sets;
- Accuracy problems: neural networks are ML algorithms that provide the state of the accuracy on many use cases. Frequently, the accuracy of the network we are building is not be satisfactory: 99% accuracy is not equal to 99% success. Legally, a 1% failure rate means not having, beyond any reasonable doubt, the certainty that the output is actually what was expected. When evaluating an ML model, it is useful to establish the so-called high bias and high variance. High bias refers to a scenario where your model "underfits" the example dataset: the model is assumed not to present a precise or representative picture of the relationship between the inputs and the predicted output. Contrarywise, high variance refers to a scenario in which the model "overfits" the dataset: it is so accurate that it is perfectly fitted to your example dataset.While seemingly a good outcome, it is a concerning one, as such models often fail to generalize to future datasets. These models might work properly for prefixed existing data, but not for general uses
- Debug problems: for a result to be demonstratable and reproducible, it is necessary to probe all steps leading to a certain result. Technically, it is difficult to accomplish a similar degree of transparency. Such criticality finds a double explanation: firstly, proceeding with real-time debugging, capable of witnessing step by step the choices made, is virtually impossible; secondly, due to the unpredictability of machine learning algorithms applied in the development of neural networks, it is not always possible to predict the variations suffered by the original mathematical model in the face of new and unknown scenarios.

The margin-of-error question becomes a matter of constitutionality, as the decision-making process must provide comprehensive and coherent reasoning from a legal and logical point of view. The need to reconstruct the logical path in a way that justifies and accounts for the results put out by the instrument clashes with the technical difficulties encountered in the process (Horsman 2019).
Justifying the results obtained requires that these instruments be used in keeping with the need to undergo authoritative measures that can be judged on the merits of their assumptions. This obstacle suggests that the use of these technologies should be limited to an auxiliary support function, of circumstantial rank, which requires the results obtained through their falsification to be evaluated at a time prior to the evaluation.
As to satisfy the gap in terms of the reliability and transparency of AI-based systems, it is essential to recognize the key role played by having a deeper and more sensitive approach to the legal reflections on the usage of digital technologies. In order to achieve this target, we strongly endorse the creation of supervised systems, those who still address interpretability to its own choices; and protecting the rights of all the parties involved in the trial, by opening up to their participation in the execution of technical operations; forging a set of certified IT skills and opening the road to the so called *humanware* in Digital Forensics field. If we do not act upon the paths of a human-centered perspective, we will not be able to take advantage from the application of AI-based systems.

### 3.3 *Nemo tenetur se detergere*

The amount of information passing every second through digital networks and devices is the preferred source of evidence in criminal proceedings: the techniques available in the field of DF for the detection of crimes and the resolution of legal cases are used on a daily bases DF

experts use a variety of technologies for the detection of crimes and the resolution of legal cases (Opijnen and Santos 2017). This obliges us to reflect, with greater consideration, on the relationship between principles and procedural rules and the new technological frontiers.

The critical profiles are highlighted above all with reference to the violation of the right to confidentiality of correspondence and privacy. The most extreme consequences of this schism develop in procedural systems based on the recognition of the right against self-incrimination, which deserve to be properly regulated.

The pervasiveness of digital investigation, due to the growth of the storage capacities, the distribution of digital services in performing daily activities over which we generate a huge amount of valuable information, and the advent of a new online reality, are now facts shared in the ordinary experience. Investigative techniques are constantly evolving and have had to undergo the transformation dictated by the entry of the digital dimension, that became a new space inside which it is possible to commit and prosecute old and new crimes. Techniques in digital investigations need to continually fit the growth and spread of computer skills in crime commission.

For this reason, they require a regulation that encourages the unfolding of skills that can safeguard the conduct of investigations in the digital field in respect of the right not to self-incriminate. It draws a distinction between the possible investigative scenarios, by setting a minimum level of warranty, such as the faculty to attend to the technical operations or a video recording that repeatable. Even creating an *ad hoc* stage in the trial to guarantee the right of a fair trial by the opening of technical schemes, such as keyword searches, is a good point to envisage a better way for the employment of those rights.

For this reason, we argue that technical and regulatory frameworks should be developed to guarantee internationally recognised fundamental rights, when they are not already established by national legislation (Saleem, Baggili, and Popov 2014). In the current scenario, increasing attention is being paid to respect for procedural guarantees in the processing of digital evidence, not only with regard to the technical requirements of admissibility but also to the limits of usability of the acquired information (Nieto et al. 2019).

On the one hand, studies aimed at raising the thresholds for the protection of the rights at stake are growing; on the other, there is a widespread reluctance to reconsider the centrality of the means of proof offered by DF techniques in ascertaining legally relevant facts (Sunde and Dror 2019; Henseler and [van Loenhout] 2018).

There are numerous attempts to save the regulatory scope of traditional institutes by adapting technological innovations to pre-existing legal concepts, rather than studying their functioning and understanding which legal rationale would be more appropriate for them. Despite the delays accumulated by legislation, there are encouraging signs of development of privacy-preserving architectures in the context of digital investigations: only the artefacts relevant to the crime being prosecuted would be exposed, while excluding any other personal information or information related to other crimes, of which one may become aware by analysing all the stored content (Opijnen and Santos 2017; Verma et al. 2019).

For these reasons, we believe that the defence of fundamental rights cannot find a justifiable compression in the availability of invasive and unregulated means.

## 4 Prospective proposals

Due to the incremental collection and sharing of Electronically Stored Information (ESI) from different sources, such as the increase and fragmentation of storage capability, the computer specialist's daily workload is evidently increased: it often requires a reactive response in a large data-set, in order to prosecute the crime and preserve the evidence.

AI/ML techniques are well suited to automate traditional tasks, possibly optimizing the time consumption and quality of the forensic process. Examples include classification of relevant evidence, detection of suspicious artefacts, recognition of suspects' faces, age calculation in child sexual exploitation material (Anda, Le-Khac, and Scanlon 2020), in addition to the creation of a framework of intelligent agents to parallelize tasks and ensure particular reliability for each of them, thanks to, for instance, privacy-preserving architecture that enables the access only case-relevant information (Verma et al. 2019).

In this context, we believe that the application of AI in DF is an appealing solution to the current and future challenges of DF, by both overcoming the limits of time shortage and ensuring reliability and admissibility of the digital evidence processed by AI forensics tools.

We also firmly believe that the human factor cannot be replaced by a machine, which is why growing a well-established *humanware* is fundamental to tackling the legal issues relating to the limits of AI in D(Casey 2017). Any digital investigator knows from their daily experience the importance of understanding how an analytic tool approaches evidence, in order to produce a reliable explanation and consequently collect admissible evidence.

This is only the first step in providing better compliance with a digital forensics framework related to the reliability of evidence, achieving reproducible results, and balancing fundamental rights with the trial's needs. The best way to tackle the previously uncovered legal issues is to cast AI in a supporting role in DF tasks.

In spite of that, how could be possibly brought out such a model? Beginning by structuring an architecture dedicated to the running of digital investigations, accessible on every prosecutors' departments. Trough the creation of a dedicated law enforcement agencies, in close interaction with the academic researchers, formed up with qualified training courses to tackle the endless evolving of DF techniques, we could probably be capable to face out the grade of ethical and legal issues caused by the introduction of AI systems into decision-making processes.

In our daily scenario we are searching, almost without any other alternative source, a digital proof even related to ancient crimes in order to find relevant artifacts that prove that

prosecuted crime. Due to this reason, we have a lack of updating regulation and building a fundamental component of a system based on the principles of a fair trial, a *humanware* fact maybe the turning point of this intricate challenge which is balancing fundamental right with the range of Digital Forensics tools based on AI potential.

For these reasons, we believe that the only sustainable solution is fighting for is to face all the ethical problems relating to AI by following a human-centred vision. In this path forward we have to raise a strong background for achieving a truly trustworthy AI ecosystem, also with the help of the EU ethics guidelines for trustworthy AI, which are focused on the development of AI-based tools that allow compliance with all laws and regulations and with ethical principles, and offering a more robust and reliable solution from both a technical and a social perspective.

This will therefore make it possible to develop technical equipment aimed at guaranteeing all of the fundamental rights that may be at risk when it comes to AI (Hamon, Junklewitz, and Sanchez 2020; Commission 2019).

***

Although this article is the result of the authors joint research, the draft (paper) has been divided as it follows: R.Brighi par.1, 3, 4; M.Ferrazzano par.2, 2.1, 2.2; L. Summa par.3.1, 3.2, 3.3.

## References


Abdelnasser Gamal, A. 2020. Artificial intelligence and humans. *International Journal of Scientific and Research Publications (IJSRP)* 10:p9970.

Anda, F.; Le-Khac, N.-A.; and Scanlon, M. 2020. Deepuage: Improving underage age estimation accuracy to aid csem investigation. *Forensic Science International: Digital Investigation* 32:300921.

Brighi, R., and Maioli, C. 2016. Un cambiamento di paradigma nelle scienze forensi. dall'armonizzazione tecnico-giuridica a una nuova cornice epistemologica. *Informatica e Diritto* XXIV:217–234.

Casey, E. 2017. The value of forensic preparedness and digital-identification expertise in smart society. *Digital Investigation* 22:1–2.

Commission, E. 2019. Ethics guidelines for trustworthy ai.

Costantini, S.; De Gasperis, G.; and Olivieri, R. 2019. Digital forensics and investigations meet artificial intelligence. *Annals of Mathematics and Artificial Intelligence* 86(1-3):193–229.

Council, N. R., et al. 2009. *Strengthening forensic science in the United States: a path forward*. National Academies Press.

Degeling, M., and Berendt, B. 2018. What is wrong about robocops as consultants? a technology-centric critique of predictive policing. *AI SOCIETY* 33:3:347 – 356.

Ferrazzano, M. 2019. *Autonomous driving e informatica forense: la prova della responsabilità in caso di sinistri*. Giappichelli.

Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; and Vayena, E. 2018. Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines* 28:689–707.

Ghafir, I.; Hammoudeh, M.; Prenosil, V.; Han, L.; Hegarty, R.; Rabie, K.; and Aparicio-Navarro, F. J. 2018. Detection of advanced persistent threat using machine-learning correlation analysis. *Future Generation Computer Systems* 89:349 – 359.

Grace, J. 2019. Machine learning technologies and their inherent human rights issues in criminal justice contexts.

Hamon, R.; Junklewitz, H.; and Sanchez, I. 2020. Robustness and explainability of artificial intelligence.

Hasan, R.; Raghav, A.; Mahmood, S.; and Hasan, M. 2011. Artificial intelligence based model for incident response. 3.

Henseler, H., and [van Loenhout], S. 2018. Educating judges, prosecutors and lawyers in the use of digital forensic experts. *Digital Investigation* 24:S76 – S82.

Horsman, G. 2019. Tool testing and reliability issues in the field of digital forensics. *Digital Investigation* 28:163–175.

King, T. C.; Aggarwal, N.; Taddeo, M.; and Floridi, L. 2020. Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics* 26:1:89–120.

Kotsoglou, K. N. 2019. Proof beyond a context-relevant doubt. a structural analysis of the standard of proof in criminal adjudication. *Artificial Intelligence and Law*.

Krivchenkov, A.; Misnevs, B.; and Pavlyuk, D. 2019. Intelligent methods in digital forensics: State of the art. *Lecture Notes in Networks and Systems* 274–284.

Lasagni, G., and Contissa, G. 2020. When it is (also) algorithms and ai that decide on criminal matters: In search for an effective remedy. *European journal of Crime, Criminal Law and Criminal Justice* 3.

Losavio, M.; Pastukov, P.; Polyakova, S.; Zhang, X.; Chow, K.; Koltay, A.; James, J. I.; and Ortiz, M. 2019. The juridical spheres for digital forensics and electronic evidence in the insecure electronic world. *Wiley Interdisciplinary Reviews: Forensic Science*.

Millar, J., and Kerr, I. 2013. Delegation, relinquishment and responsibility: The prospect of expert robots. *SSRN Electronic Journal*.

Mitchell, F. 2014. The use of artificial intelligence in digital forensics: An introduction. *Digital Evidence and Electronic Signature Law Review* 7.

Nieto, A.; Rios, R.; Lopez, J.; Ren, W.; Wang, L.; Choo, K.-K. R.; and Xhafa, F. 2019. *Privacy-aware digital forensics*.

Opijnen, M., and Santos, C. 2017. On the concept of relevance in legal information retrieval. *Artificial Intelligence and Law* 25:65–87.

Quattrocolo, S.; Anglano, C.; Canonico, M.; and Guazzone, M. 2020. *Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings*, volume 20.

Russell, S., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*. USA: Prentice Hall Press, 3rd edition.

Saleem, S.; Baggili, I.; and Popov, O. 2014. Quantifying relevance of mobile digital evidence as they relate to case types: A survey and a guide for best practices. *The Journal of Digital Forensics, Security and Law* 9.

Sanchez, L.; Grajeda Mendez, C.; Baggili, I.; and Hall, C. 2019. A practitioner survey exploring the value of forensic tools, ai, filtering, safer presentation for investigating child sexual abuse material (csam). *Digital Investigation* 29:S124–S142.

Sanger, R. 2018. Forensics: Educating the lawyers. *SSRN Electronic Journal*.

Sartor, G. 1992. Artificial intelligence in law and legal theory. *Current Legal theory* 10:1–59.

Sartor, G. 1998. Judicial applications of artificial intelligence. *Artificial Intelligence and Law* 7:157–372.

Sommaggio, P., and Marchiori, S. 2020. Moral dilemmas in the a.i. era: A new approach.

Sunde, N., and Dror, I. 2019. Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. *Digital Investigation* 29.

Trifonov, R.; Yoshinov, R.; Manolov, S.; Tsochev, G.; and Pavlova, G. 2019. Artificial intelligence methods suitable for incident handling automation. *MATEC Web of Conferences* 292:01044.

Verma, R.; Govindaraj Dr, J.; Chhabra, S.; and Gupta, G. 2019. Df 2.0: An automated, privacy preserving, and efficient digital forensic framework that leverages machine learning for evidence prediction and privacy evaluation. *Journal of Digital Forensics, Security and Law* 14(2):3.

Vuille, J.; Lupària, L.; and Taroni, F. 2017. Scientific evidence and the right to a fair trial under article 6 echr. *Law, Probability and Risk* 16(1):55–68.